

WHAT IS CLAIMED IS:

1. A method that analyzes mass spectra using a digital computer, the method comprising:
 - 5 a) entering into a digital computer a data set obtained from mass spectra from a plurality of samples, wherein each sample is, or is to be assigned to a class within a class set comprising two or more classes, each class characterized by a different biological status, and wherein each mass spectrum comprises data representing signal strength as a function of time-of-flight, mass-to-charge ratio, or a
10 value derived from time-of-flight or mass-to-charge ratio; and
 - b) forming a classification model which discriminates between the classes in the class set, wherein forming comprises analyzing the data set by executing code that embodies a classification process comprising a recursive partitioning process.
- 15 2. The method of claim 1 wherein the mass spectra are selected from the group consisting of MALDI spectra, surface enhanced laser desorption/ionization spectra, and electrospray ionization spectra.
3. The method of claim 1 wherein the class set consists of exactly two classes.
20
4. The method of claim 1 wherein the samples comprise biomolecules selected from the group consisting of polypeptides and nucleic acids.
5. The method of claim 1 wherein the samples are derived from a eukaryote, a
25 prokaryote or a virus.
6. The method of claim 1 wherein the different biological statuses comprise a normal status and a pathological status.
- 30 7. The method of claim 1 where the different biological statuses comprise un-diseased, low grade cancer and high grade cancer.
8. The method of claim 1 wherein the different biological statuses comprise a drug treated state and a non-drug treated state.

9. The method of claim 1 wherein the different biological statuses comprise a drug-responder state and a drug-non-responder state.
- 5 10. The method of claim 1 wherein the different biological statuses comprise a toxic state and a non-toxic state.
11. The method of claim 10 wherein the toxic state results from exposure to a drug.
- 10 12. The method of claim 1 wherein the data set is a known data set, and each sample is assigned to one of the classes before the data set is entered into the digital computer.
- 15 13. The method of claim 1 wherein forming the classification model comprises using pre-existing marker data to form the classification model.
14. The method of claim 1 wherein the data set is formed by:
- 20 detecting signals in the mass spectra, each mass spectrum comprising data representing signal strength as a function of mass-to-charge ratio;
- clustering the signals having similar mass-to-charge ratios into signal clusters;
- selecting signal clusters having at least a predetermined number of signals with signal intensities above a predetermined value;
- 25 identifying the mass-to-charge ratios corresponding to the selected signal clusters; and
- forming the data set using signal intensities at the identified mass-to-charge ratios.
- 30 15. The method of claim 1 wherein forming the classification model comprises at least one of identifying features that discriminate between the different biological statuses, and learning.

16. The method of claim 1 wherein the classification process is a binary recursive partitioning process.
17. The method of claim 1 further comprising:
- 5 c) interrogating the classification model to determine if one or more features discriminate between the different biological statuses.
18. The method of claim 1 further comprising:
- c) repeating a) and b) using a larger plurality of samples.
- 10 19. The method of claim 1 wherein the classification process is a classification and regression tree process.
20. The method of claim 1 further comprising forming the data set, wherein
- 15 forming the data set comprises obtaining raw data from the mass spectra and then preprocessing the raw mass spectra data to form the data set.
21. The method of claim 1 wherein the different classes are selected from exposure to a drug, exposure to one of a class of drugs and lack of exposure to a drug
- 20 or one of a class of drugs.
22. The method of claim 1 wherein the each mass spectrum comprises data representing signal strength as a function mass-to-charge ratio or a value derived from mass-to-charge ratio.
- 25 23. A method for classifying an unknown sample into a class characterized by a biological status using a digital computer, the method comprising:
- a) entering data obtained from a mass spectrum of the unknown sample into a digital computer; and
- 30 b) processing the mass spectrum data using the classification model formed by the method of claim 1 to classify the unknown sample in a class characterized by a biological status.
23. The method of claim 23 wherein the class is characterized by a disease status.

24. The method of claim 23 wherein the different biological statuses comprise un-diseased, low grade cancer and high grade cancer.
- 5 25. The method of claim 23 wherein the class is characterized by exposure to a drug of one of a class of drugs.
26. The method of claim 23 wherein the class is characterized by response to a drug.
- 10 27. The method of claim 23 wherein the class is characterized by a toxicity status.
28. A method for estimating the likelihood that an unknown sample is accurately classified as belonging to a class characterized by a biological status using a digital computer, the method comprising:
- 15 a) entering data obtained from a mass spectrum of the unknown sample into a digital computer; and
- b) processing the mass spectrum data using the classification model formed by the method of claim 1 to estimate the likelihood that the unknown sample
- 20 is accurately classified into a class characterized by a biological status.
29. A computer readable medium comprising:
- a) code for entering data obtained from a mass spectrum of an unknown sample into a digital computer; and
- 25 b) code for processing the mass spectrum data using the classification model formed by the method of claim 1 to classify the unknown sample in a class characterized by a biological status.
30. A system comprising:
- 30 a gas phase ion spectrometer;
- a digital computer adapted to process data from the gas phase ion spectrometer; and
- the computer readable medium of claim 29 in operative association with the digital computer.

31. The system of claim 30 wherein the gas phase ion spectrometer is adapted to perform a laser desorption ionization process.
- 5 32. A computer readable medium comprising:
- a) code for entering data obtained from a mass spectrum of an unknown sample into a digital computer; and
 - b) code for processing the mass spectrum data using the classification model formed by the method of claim 1 to estimate the likelihood that the unknown
- 10 sample is accurately classified into a class characterized by a biological status.
33. A system comprising:
- a gas phase ion spectrometer;
 - a digital computer adapted to process data from the gas phase ion
- 15 spectrometer; and
- the computer readable medium of claim 32 in operative association with the digital computer.
34. The system of claim 33 wherein the gas phase ion spectrometer is adapted to
- 20 perform a laser desorption ionization process.
35. A computer readable medium comprising:
- a) code for entering data derived from mass spectra from a plurality of samples, wherein each sample is, or is to be assigned to a class within a class set of
- 25 two or more classes, each class characterized by a different biological status, and wherein each mass spectrum comprises data representing signal strength as a function of time-of-flight, mass-to-charge ratio or a value derived from mass-to-charge ratio or time-of-flight; and
- b) code for forming a classification model using a classification process,
- 30 the classification process comprising a recursive partitioning process, wherein the classification model discriminates between the classes in the class set.
36. The computer readable medium of claim 35 wherein the classification process is a classification and regression tree process.

37. A system comprising:
a gas phase ion spectrometer;
a digital computer adapted to process data from the gas phase ion
5 spectrometer; and
the computer readable medium of claim 35 in operative association with the
digital computer.
38. The system of claim 37 wherein the gas phase ion spectrometer is adapted to
10 perform a laser desorption ionization process.